Effect of System Latency in Dynamic Virtual Acoustic Environments

Elizabeth M. Wenzel

The aviation environment contains multiple channels of auditory and visual information that must be accessed under high-stress, high-workload conditions. Auditory displays that controllers and pilots currently use can be significantly enhanced by utilizing three-dimensional (3-D) audio technology. Separate channels of auditory information can be placed at different virtual locations (1) to enhance situation awareness (e.g., airborne or ground traffic collision avoidance alerts, taxiway navigation aids and announcements); (2) to increase intelligibility (through the use of binaural delivery systems); and (3) to reduce auditory fatigue. Enabling these displays requires the development of specialized hardware systems for rendering virtual audio, the assessment of their engineering performance characteristics, and the perceptual validation of the spatial cues rendered by such systems.

Previous research suggests that auditory localization errors are minimized when a virtual acoustic environment (VAE) is dynamic, that is, when listeners are allowed to move their heads and the spatial cues change appropriately in real time. In such a display, knowledge of system parameters such as latency is critical for assessing real-time performance, and it is important that these parameters be carefully defined and measured. Psychoacoustic parameters such as the minimum audible movement angle (MAMA) can then be used as target guidelines to assess whether a given system meets perceptual requirements. In addition, such measurements enable systematic perceptual studies of the effect of degrading system latency on localization accuracy.

In a VAE, the total system latency (TSL) refers to the time elapsed from the initiation of an event or action, such as a movement of the head, until the consequences of that action cause the equivalent change in the virtual sound source location. Latencies are contributed by individual components of a VAE system, including tracking devices, signal processors, software to control these devices, and communications lines. There is no reason to expect that a system's latency remains constant over time. Thus, measurements of the mean, standard deviation, and range of the TSL provide a better characterization of this parameter.

This article reports on measurements of TSL for the virtual audio system used in previous studies of localization with and without head motion. The system consisted of a Convolvotron spatialization device that simulates direct-path spatial cues using 256-point, minimum-phase head-related transfer functions (HRTFs). It received head-position data from a Polhemus Fastrak (40-hertz update rate). In order to measure latency, a special HRTF map was constructed, which contained a single impulse at one map location and zeroes at all other locations. Latency measurements were conducted using the testbed shown in the first figure. The Fastrak receiver was mounted on the end of a mechanical swing arm with an optical switch that detected when the arm passed through a preset threshold position. This event threshold was considered analogous to the initiation of head (or source) motion in a VAE and began the universal counter's TSL timing cycle. At the same time, the tracker sent data to the Convolvotron in polled mode via a serial line, and a signal generator fed a 6000-hertz square wave to one of the input channels. Before the threshold was crossed, the Convolvotron was set to a zeroed map location so that no signal passed through the output channel. The experimenter then pushed the swing arm through the threshold position. The next tracker data sample that was received after threshold crossing caused the Convolvotron to switch to the nonzero map location. The square wave then passed through the system and terminated the timing cycle of the universal counter whose reading was considered to be the TSL. A total of 117 such measurements were taken: mean = 54.3, standard deviation = 8.8, range = 35.4 to 74.6 milliseconds.

Examination of the head motions that listeners use to aid localization suggests that the angular velocity of some head motions (in particular, left-right yaw) may be as fast as about 175 degrees per second for short time periods (e.g., about 1 second). A maximum TSL of 75 milliseconds could potentially result in short-term undersampling of relative listener-source motion as well as positional instability of the simulated source. For example, in the second figure, head motion yaw for an individual listener is plotted

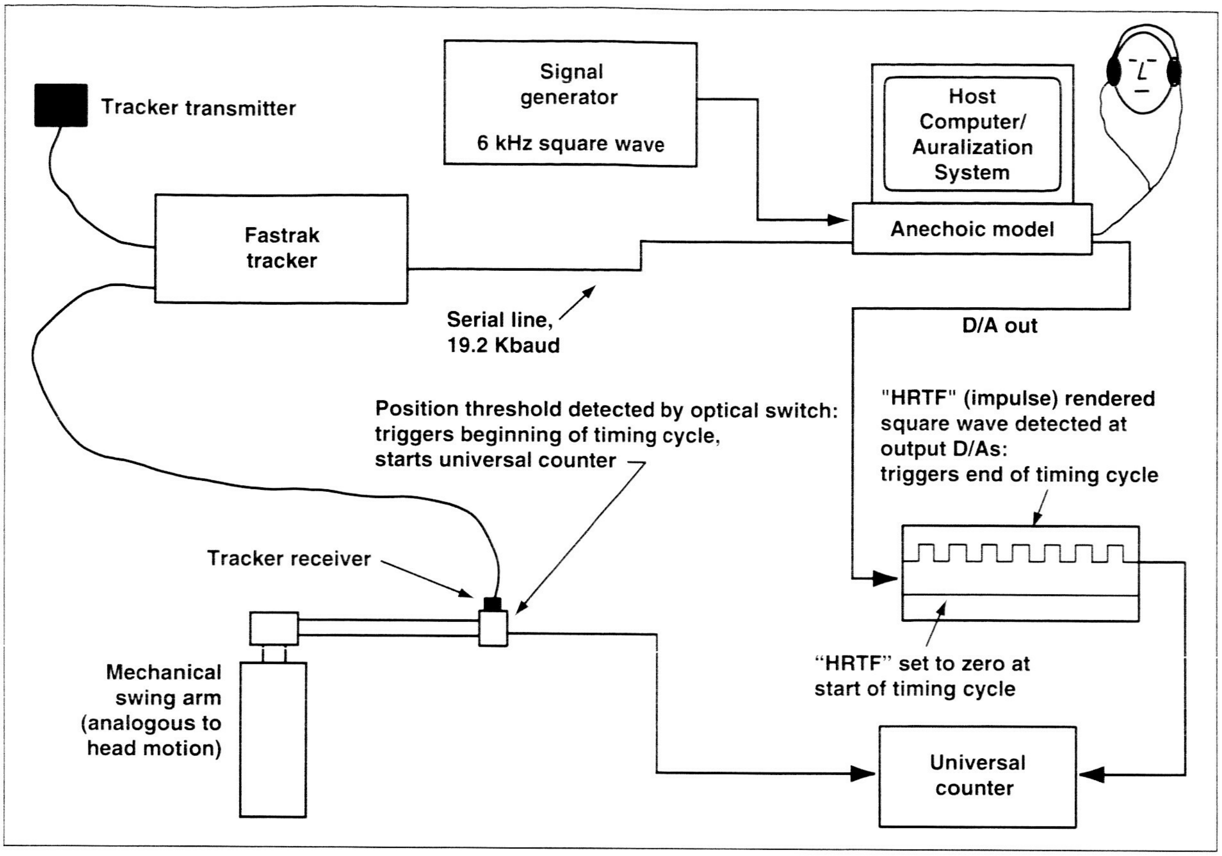

Fig. 1. Testbed for measuring total system latency.

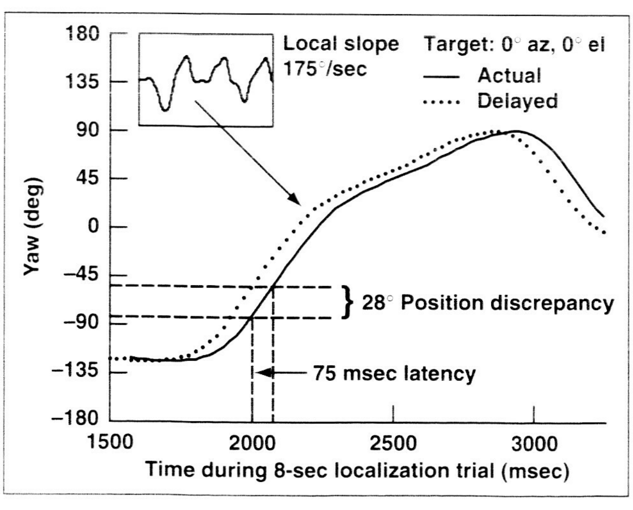

Fig. 2. Illustration of position displacement caused by latency.

as a function of time (undelayed and delayed by 75 milliseconds) during localization of a virtual source. The inset shows the entire 8-second trial. In regions of the head-motion trace where angular head motion is large (local slope = 175 degrees per second), a TSL of 75 milliseconds could result in a relative position discrepancy of about 28 degrees between actual head orientation and the rendered direction of the source. From psychophysical studies of the MAMA for real, moving sound sources, one can infer that the minimum perceptible TSL for a virtual audio system should be no more than about 69 milliseconds for an angular source velocity of 180 degrees per second. Thus, the positional displacement of the simulated source caused by TSL may have occasionally exceeded the perceptible threshold. In fact, listeners did not report any obvious instability in source position.

Point of Contact: E. Wenzel
(650) 604-6290
bwenzel@mail.arc.nasa.gov